FORECASTING RESEARCH DIVISION
Technical Report No. 67

An operational trial of Synoptic-dependent Background Errors

by
Colin Parrett
October 1993

Forecasting Research

Technical Report No. 67

An operational trial of Synoptic-dependent Background Errors

by

Colin Parrett

October 1993

Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
England

N.B. This paper has not been published. Permission to quote from it must be obtained from the above division of the Met. Office.

1. Introduction

A Parallel Suite trial of synoptic-dependent background errors (BGEs) for upper level wind and temperature data was held from 22nd July to 2nd August 1993 [Parallel Suite experiment 12]. These BGEs, dependent on meteorological parameters such as wind speed and temperature gradient, replaced the climatological BGEs (obtained from the previous Cyber model's statistics) used operationally in the quality control and assimilation. [Details of the synoptic-dependent BGEs can be found in Parrett, 1992.] The trial also included some associated tuning of the observation errors, quality control and assimilation parameters (see Appendix A for details).

A continuous assimilation cycle was run, with one daily forecast to T+120 from the 12Z parallel run. The forecast from 12Z 26th July was not run, because of hardware problems. A parallel observation processing database (OPD) was produced with just one run missing, although there were problems with the operational aircraft OPD during the trial period, with 3 consecutive days of data missing. Some charts were produced during the trial to show the differences between the trial and operational runs; and after the trial some of the (archived) fields-files and OPDs were transferred to a workstation for more detailed investigations.

This report is divided into 3 main sections. Section 2 below considers the impact of the synoptic-dependent BGE runs on the quality control of data. Section 3 contains objective results: the trial analyses and forecasts have been verified objectively against observations using the same system as used to verify the operational runs, and a comparison of these results is presented. Some observation-background (o-b) results from the OPD are also shown in section 3. Section 4 contains a subjective assessment of the last 5 days of the trial, a period free from any OPD/fields-file archiving problems, which has been looked at in some detail. Finally, section 5 contains a brief summary of the trial results and some concluding remarks.

2. Effect on quality control of data

The different BGEs in the trial and operational runs (see Figures 1 and 2 for examples of 250hPa wind component BGEs and 500hPa temperature BGEs, respectively), and the different T+6 backgrounds from the assimilation cycles, led to some different quality control decisions. Fewer radiosonde winds were flagged at all levels above the boundary layer, and Figure 3(a) shows that the percentage of winds flagged was reasonably constant with level in the trial run, whereas operationally a much larger percentage was flagged near jet levels - we have no reason to believe that in general the wind reports from these levels are less accurate than those below and above. The most likely reason for more flagging operationally, is that where the winds are stronger the o-b differences are greater, which is not reflected in the operational BGEs. There was also a reduction in the number of radiosonde temperatures flagged, particularly in the stratosphere (Figure 3(b)). Overall, out of ~21000 radiosonde reports, only 288 winds (0.11%) and 653 temperatures (0.31%) were flagged in the trial runs, compared to 636 winds (0.24%) and 843 temperatures (0.40%) in the operational runs. There were also slightly fewer aircraft reports flagged in the trial: 0.28% of winds compared with 0.30% operationally, and 0.83% of temperatures compared with 0.85% operationally. Slightly more satellite cloud-track winds were flagged in the trial, 2.66% compared with 2.45% operationally.

The different quality control decisions for radiosondes and aircraft reports on 4 days of the final week of the trial were investigated further. There were 77 different decisions for wind in these runs - 67 were flagged operationally but passed in the trial (61 sondes and 6 aircraft), and 10 passed operationally but were flagged in the trial (3 sondes and 7 aircraft). There were fewer different quality control decisions for temperature reports, mostly in the stratosphere, and these have not been looked at in detail. Most of the extra observations in the trial runs were single sonde levels (or aircraft reports) that were just over the o-b limit for flagging operationally, but were brought within the limit in the trial. No obvious gross errors were found in the extra data. Figures 4 and 5 show examples in which the larger BGEs in the trial run (Figure 1(a) cf Figure 1(b)) gave larger total error variances,

relaxing the thresholds for flagging and allowing the two observations concerned (circled) to pass the quality control checks; whereas they were flagged operationally [see Ingleby and Parrett, 1993a, for details about the quality control scheme]. The effect on the wind analysis at 250 hPa of the extra level of observation 25563 (Figure 4), was to increase the jet strength from 83 knots operationally to 91 knots in the trial (the 12 knot difference at the observation location was the largest difference between analyses seen in the N.Hemisphere). The single level of observation 94302 (in north west Australia) that was flagged operationally had less impact on the trial analysis, although the jetstream was extended upstream and broadened slightly (Figure 5(a) cf Figure 5(b)).

Figures 6(a) and 6(b) show examples of aircraft wind reports (circled) that were flagged in the trial run, whereas they passed the quality control checks operationally. In these two cases (from near Hawaii and southeast of Japan, respectively) the synoptic-dependent BGE was smaller than the operational BGE (e.g. $4.6ms^{-1}$ cf $6.3ms^{-1}$ in Figure 6(a)), contributing to the reports being flagged in the trial run. Comparing the flagged reports with surrounding observations in Figures 6(a) and 6(b), the trial run seems to have been correct in flagging them in both cases. For the case in Figure 6(a), another co-located observation with the same wind direction but slightly stronger speed was flagged in both runs, supporting the idea that this observation was correctly flagged in the trial run.

3. Objective verification

3.1 Assimilation cycle diagnostics

Table 1 below shows that the differences (trial - operational) in the global rms o-b values are generally very small. Where there are differences they may in part be due to the different quality control decisions causing some differences in the data used in the 2 runs - particularly for the model level 12 vector wind, for which there were 49 more observations in the trial runs (presumably having quite large o-b values to be flagged in the operational run).

Table 1. Global root-mean-square observation-background values

Data type, level	Trial o-b	Operational o-b
Ship Pmsl	1.45 hPa	1.45 hPa
Synop Pmsl	1.31 hPa	1.30 hPa
Sonde T, lev 4 (846 hPa)	1.39 °C	1.39 °C
Sonde T, lev 8 (491 hPa)	1.12 °C	1.12 °C
Sonde T, lev 12 (246 hPa)	1.58 °C	1.57 °C
Sonde V, lev 4 (846 hPa)	4.16 m/s	4.16 m/s
Sonde V, lev 12 (246 hPa)	6.29 m/s	6.19 m/s
Sonde RH, lev 4 (846 hPa)	14.1 %	14.1 %
Airep V, above 400 hPa	5.94 m/s	5.99 m/s

Figure 7 gives a breakdown by wind speed category of the rms o-b for vector wind for radiosondes north of 30°N. For this comparison the same set of sondes/levels was used from both runs - those passing the quality control checks in both the trial and operational runs. At low wind speeds the trial's rms o-b was slightly higher, but at the more important higher speed ranges the trial's values were slightly lower (better). The higher speed ranges are more important, because that is where the largest differences from background are found, and any reduction is welcome. The global rms o-b values for all sondes, at all levels that were in both the trial and operational runs, were 5.06ms^{-1} operationally and 5.07ms^{-1} in the trial; so the trial was just 0.2% worse overall, mostly due to the contribution from the low wind speed ranges.

It is encouraging to note from Table 1 that for aircraft winds there was an improvement globally of nearly 1% in the trial. Table 2 below contains some results produced by R W Lunnon (pers. comm.),

Table 2. ASDAR Integrated Headwind Components over the N.Atlantic

		Analysis		Background	
		Oper	Trial	Oper	Trial
RMS diff from ASDAR:		2.86	2.78 (+2.8%)	4.03	3.91 (+3.0%)
Std Devn:	33.20 (ASDAR)	32.75	32.82	31.78	31.92

which show a substantial 3% improvement in the trial for rms o-b of ASDAR headwinds over the North Atlantic. Also shown, are the

standard deviations of the integrated headwind components, which should be close to the 33.2 knots observed. The trial run is also better here, being slightly closer in both analyses and backgrounds.

3.2 Verification of analyses and forecasts

Table 3 below shows the percentage change (operational/trial) in rms fit to radiosonde observations at 4 forecast times and for 4 different verification areas (area 2 ≡ N.Atlantic sector, area 200 ≡ north of 30°N, area 300 ≡ 30°N to 30°S, area 400 ≡ south of 30°S). There was generally a modest improvement in the fit to data at T+0 for the trial runs, especially 250 hPa wind. But by T+24 and beyond there was essentially a null impact, with some variables/times/areas slightly better and others slightly worse. [N.B. There is more

Table 3. Percentage change in rms fit to data (+ve => trial better)

Variable	Area	T+0	T+24	T+48	T+120
850 hPa height	2	-1.2	-1.1	**-2.0**	0.1
	200	-0.3	-0.1	-0.9	-1.2
	300	-0.5	-1.4	-1.9	-1.6
	400	1.4	-0.7	1.9	**3.9**
500 hPa height	2	0.6	-1.2	-0.5	**-2.8**
	200	0.7	0.6	0.0	-1.5
	300	0.8	0.1	-0.8	1.1
	400	**-2.6**	-0.7	1.7	**2.2**
250 hPa height	2	0.4	0.3	-0.5	**-2.4**
	200	0.2	0.6	0.1	-1.3
	300	0.8	0.3	-0.5	0.8
	400	-0.5	-1.6	0.2	**2.8**
850 hPa temp.	2	1.8	0.7	0.6	-0.3
	200	1.2	0.0	-0.2	-0.7
	300	**2.7**	0.2	0.7	**3.2**
	400	**2.3**	-1.8	-0.4	-1.6
850 hPa wind	2	0.7	0.0	-0.5	-0.7
	200	-0.1	0.3	-0.3	-0.8
	300	0.0	-0.3	0.1	0.8
	400	-1.1	-0.2	-0.3	-1.2
500 hPa wind	2	1.8	-0.1	-1.3	-1.9
	200	1.2	0.1	-0.6	-1.1
	300	0.2	0.0	0.1	-0.1
	400	**3.6**	-0.1	-0.5	**-3.0**
250 hPa wind	2	**2.9**	-0.2	-0.4	-1.4
	200	**2.8**	-0.4	-0.4	-0.5
	300	**2.0**	1.1	0.5	0.6
	400	**2.6**	0.5	**2.1**	**-4.8**

variability in the scores for area 400, because of the limited data

coverage in the Southern Hemisphere. There are also some inconsistencies between forecast times, because of different numbers of forecasts verified - see Appendix B for some suggested improvements to the Parallel Suite verification.]

4. Subjective assessment

A subjective assessment was carried out, concentrating on the second half of the trial period. This avoided the problems with the fields-file archiving and the OPD (prior to 28th July), and allowed a 'spin-up' period. The differences between the trial and operational forecasts for the northern hemisphere were quite small - even the forecasts from the analyses shown in Figure 4 (with the 12 knot difference) showed only small differences at T+24 (2-3 knots). Looking at the jets over the north Atlantic sector, there was little to choose between the 2 runs, with some jets being forecast slightly better in the trial runs and some slightly better in the operational runs. Figures 8 and 9 show a typical example of differences in the 250 hPa wind forecasts, from 12Z 29th July. At T+24 (Figure 8) there was a 10 knot difference S.E. of Greenland (stemming from a 3 knot difference in the analyses on the 29th), with the jet slightly better forecast in the trial run; but the differences between the forecasts was small compared to the differences between the verifying analysis and the forecasts. By T+48 (Figure 9) the differences in the N.Atlantic have decreased, with other differences appearing. The trial run seems to be better near the Brest peninsular and south west England, while the differences over Canada favour both the trial and the operational run in different places; but again both forecasts of jet strength differ markedly from the verifying analysis.

The largest differences between the trial and operational runs were over the oceans south of ~45°S, where it is difficult to verify which run was best, because the trial and operational analyses were often markedly different. In the region of Australasia, where the analyses were in better agreement, the differences between the forecasts were also generally smaller. Again there was no consistent picture of either the trial or operational runs being better than the other, just a few features being forecast slightly better in each. Figure 10 shows an example of a 3-day forecast of P_{msl} with

(trial-oper) differences in two areas where the verifying analyses agreed. The trough between Australia and New Zealand is better forecast in the trial run; whereas the low east of S.America is slightly better positioned in the operational run, but its central pressure is better in the trial run. The other differences are difficult to verify, because of lack of data and differences in the verifying analyses. Figure 11 shows an example of typical differences in the thickness and height fields at T+48, with some differences in areas where the trial and operational analyses agree and can be used to verify the forecasts. The warm ridge over New Zealand is slightly better forecast in the trial run. The baroclinic zone near southern Chile is overdone in both forecasts, but the trial is generally slightly closer to the analysis.

For the last 5 days of the trial, the trial and operational analyses of jets over the N.Atlantic, eastern N.America and W.Europe were verified subjectively against observations of wind from aircraft and sondes. Out of 28 jets looked at, 14 were analysed slightly better in the trial, compared with only 3 analysed better operationally (11 were analysed the same). The average speed of the jet maxima (at 250 hPa) was slightly higher in the trial at 84.9 knots, compared to 84.3 knots operationally, but still much lower than observed (90.6 knots).

5. Conclusion

The synoptic-dependent BGEs in the trial were often larger than the operational BGEs in regions of strong winds and large temperature gradients, and this led to fewer radiosonde and aircraft reports being flagged by the quality control in the trial runs. There was also a general improvement in the fit to data at T+0, resulting from the inclusion of these extra observations (which generally seemed to be good) and the greater weight given to the data in regions with high BGEs. This resulted in better T+6 forecasts, as shown by the improved OPD statistics for aircraft winds and for the stronger radiosonde winds.

There was a null impact overall in the objective verification of the forecasts for 1-5 days ahead. Although, with the improved quality control decisions with the synoptic-dependent BGEs, there

remains the potential for a positive forecast impact where the data is in a critical development area. Following the encouraging results from this trial, the decision was taken to introduce the synoptic-dependent BGEs operationally in November 1993.

References

Bell RS, Lorenc AC, Macpherson B, Swinbank R and Andrews P, 1993: The Analysis Correction Data Assimilation Scheme. *Unified Model Documentation Paper 30, version 4.*

Bell RS, Macpherson B, Robinson D and Andrews P, 1993: Technical details of the UM Data Assimilation scheme. *UMDP P3.*

Ingleby NB and Parrett CA, 1993a: Quality control of atmospheric data. *UMDP 32, version 2.*

Ingleby NB and Parrett CA, 1993b: Quality control of atmospheric data - technical documentation. *UMDP P52, version 2.*

Parrett CA, 1992: Background errors for the quality control and assimilation of atmospheric observations in the Unified Model - the situation in July 1992. *Forecasting Research Technical Report No.22.*

Appendix A. Tuning of the observation errors, quality control and assimilation

Two case studies were carried out with the synoptic-dependent BGEs, both consisting of 54 hours of assimilation with four 12-hour forecasts (one case from July 1992 and one from December 1992). As a consequence of these detailed studies, the following tuning measures were decided upon.

(a) The observation errors for wind for radiosonde stations with good wind-finding systems and for aircraft reports were reduced by 10-15%. This is in line with observation-background (o-b) statistics presented by Parrett (1992) and was designed to reduce the error ratio (obs error / BGE) for winds. This was to bring the error ratios more into line with present operational values and reduce the need for retuning the assimilation to fit the data.

(b) The initial probability of gross error (PGE) for radiosonde winds was reduced from 1.5% to 0.4% (which is approximately the percentage flagged operationally at the time of the case studies), and for aircraft winds the PGE was reduced from 3.0% to 1.5%.

(c) The "tophat" function for the distribution of flagged winds was reduced from 2×10^{-4} to 10^{-5} for radiosondes, and reduced from 10^{-4} to 5×10^{-5} for aircraft reports.

[N.B. The above package of changes (a-c) slightly increases the cut-off o-b value for flagging of typical 300mb winds: from 22.3ms^{-1} operationally to 23.1ms^{-1} [see Ingleby and Parrett (1993a) for other examples and formulae used]. Forecasters in the Central Forecast Office (CFO) have complained in the past that the flagging of sonde winds is too strict operationally, so this is a small step in the right direction.]

(d) The assimilation was also tuned slightly, in order to minimise the changes in the weights given to the data in the runs with synoptic-dependent BGEs. Changes were made to the nudging coefficients (UMDP P3 and UMDP 30), as shown in Table A1.

Table A1. Nudging coefficients ($*10^{-4}$)

	(a) Used operationally					(b) Used in Trial				
	P	T	V	V_{10m}	RH	P	T	V	V_{10m}	RH
NH	5.0	5.0	6.0	6.0	5.0	5.0	5.0	6.6	6.3	5.0
TR	3.0	3.0	3.5	3.5	3.0	3.0	3.0	4.0	3.8	3.0
SH	3.5	3.5	4.0	4.0	3.5	3.8	3.8	4.3	4.0	3.5

Appendix B. Suggestions for improvements to the Parallel Suite

There are several ways in which the objective verification of the analyses and forecasts in the Parallel Suite could be improved somewhat. Currently, the objective verification stops on the day that each trial finishes, which meant that for the BGE trial the last five 5-day forecasts (and last four 4-day forecasts, etc.) were not included in the verification. So, for the 12-day trial, the maximum number of 5-day forecasts included in the verification was 7, whereas the maximum number of 1-day forecasts included in the verification was 11. This leads to inconsistencies in the verification for different forecast times. Also, it is rather a waste of resources, because there is no point in running forecasts if they cannot be verified.

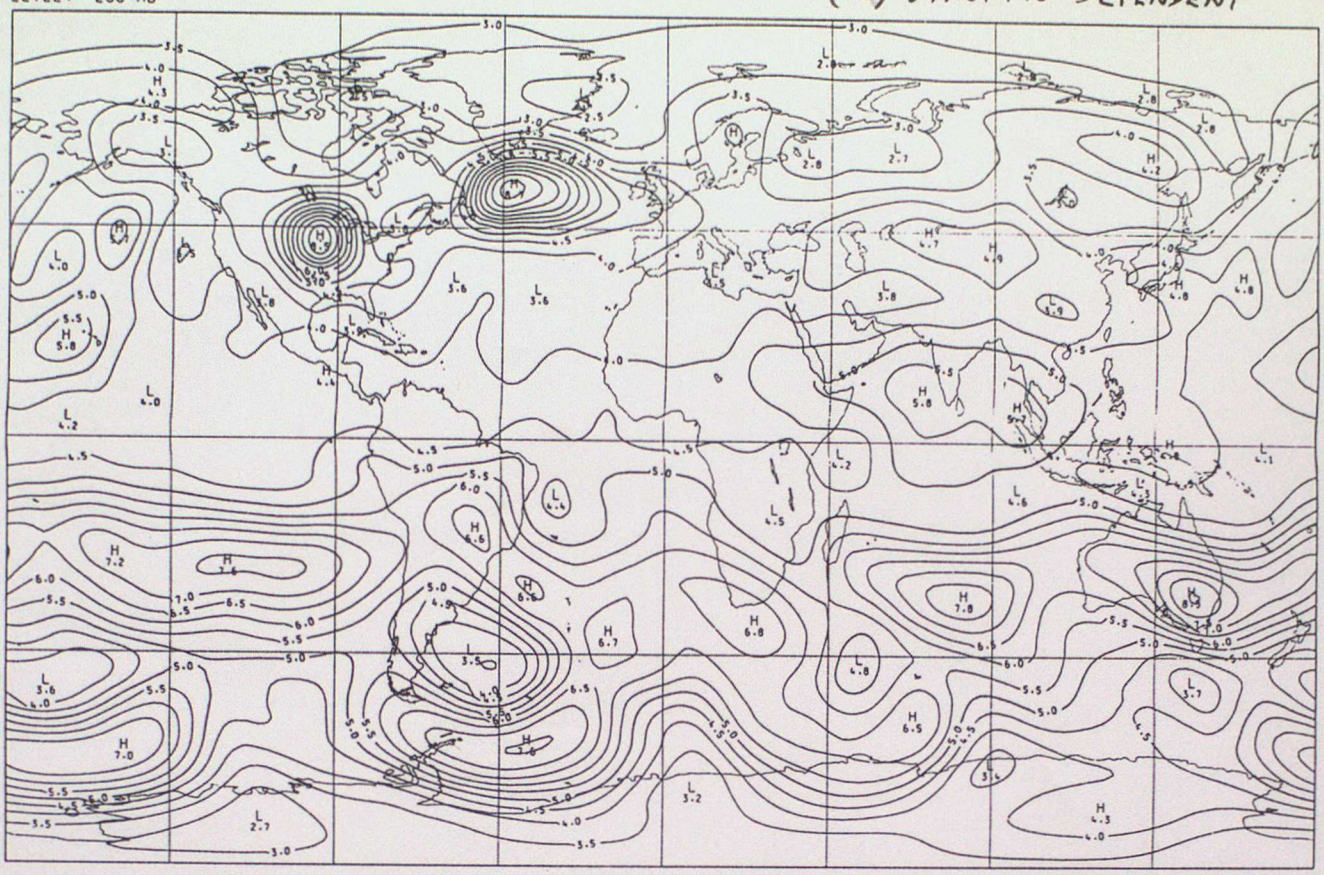
It would be better to verify both trial and operational runs against the operational observation files. Then it would be possible (perhaps?) to continue the verification beyond the 'end' of the trial. Using the operational observation files would also avoid inconsistencies in the verification comparison caused by using different observations for the trial and operational verification.

The objective verification of forecasts against analyses has been ignored in the BGE trial, because of differences in the trial and operational analyses. Currently, the trial forecasts are verified against the trial analyses and the operational forecasts are verified against the operational analyses. For a valid comparison of the verification scores, both trial and operational forecasts should be verified against the operational analyses and also against the trial analyses. Then, if the signal from the verification against operational analyses agrees with the signal from the verification against trial analyses, confidence can be placed in these results.

Figure 1.

BACKGROUND ERROR FIELD (STANDARD DEVIATION)
FOR WIND COMPONENT (M/S)
VALID AT 12Z ON 1/ 8/1993
LEVEL: 200 MB

(a) SYNOPTIC-DEPENDENT



(b) CLIMATOLOGICAL

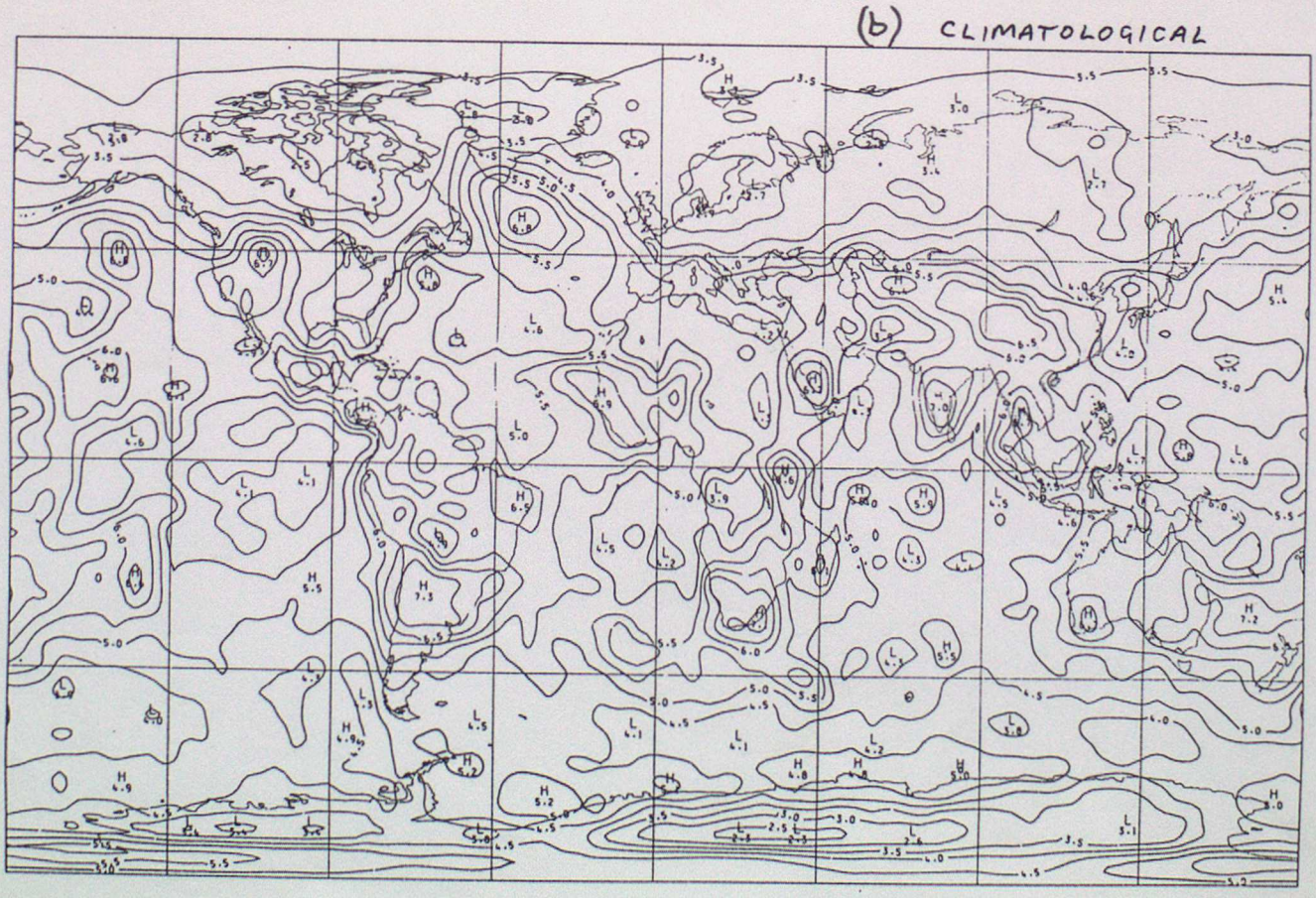
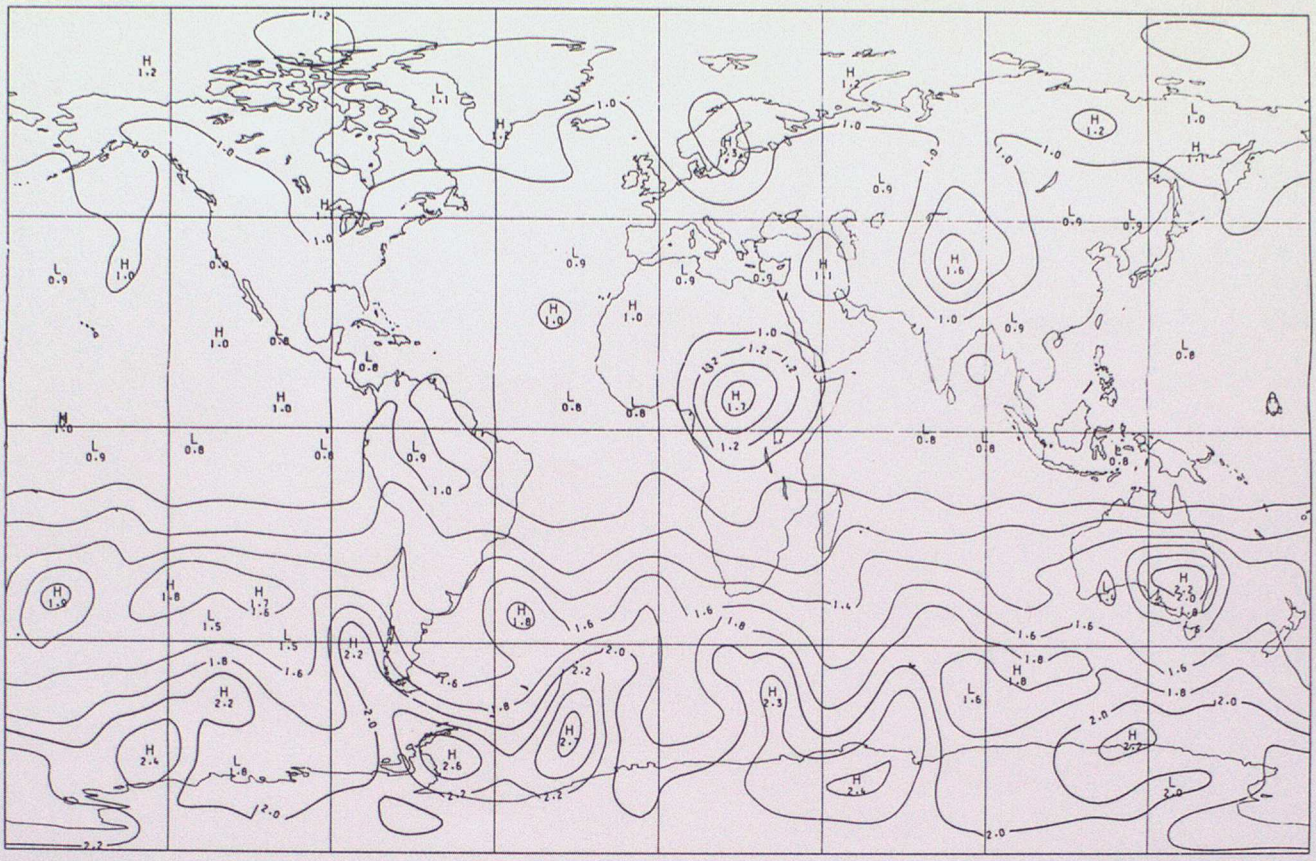


Figure 2.

BACKGROUND ERROR FIELD (STANDARD DEVIATION)
FOR TEMPERATURE (DEG C)
VALID AT 12Z ON 1/ 8/1993
LEVEL: 500 MB

(a) SYNOPTIC-DEPENDENT



(b) CLIMATOLOGICAL

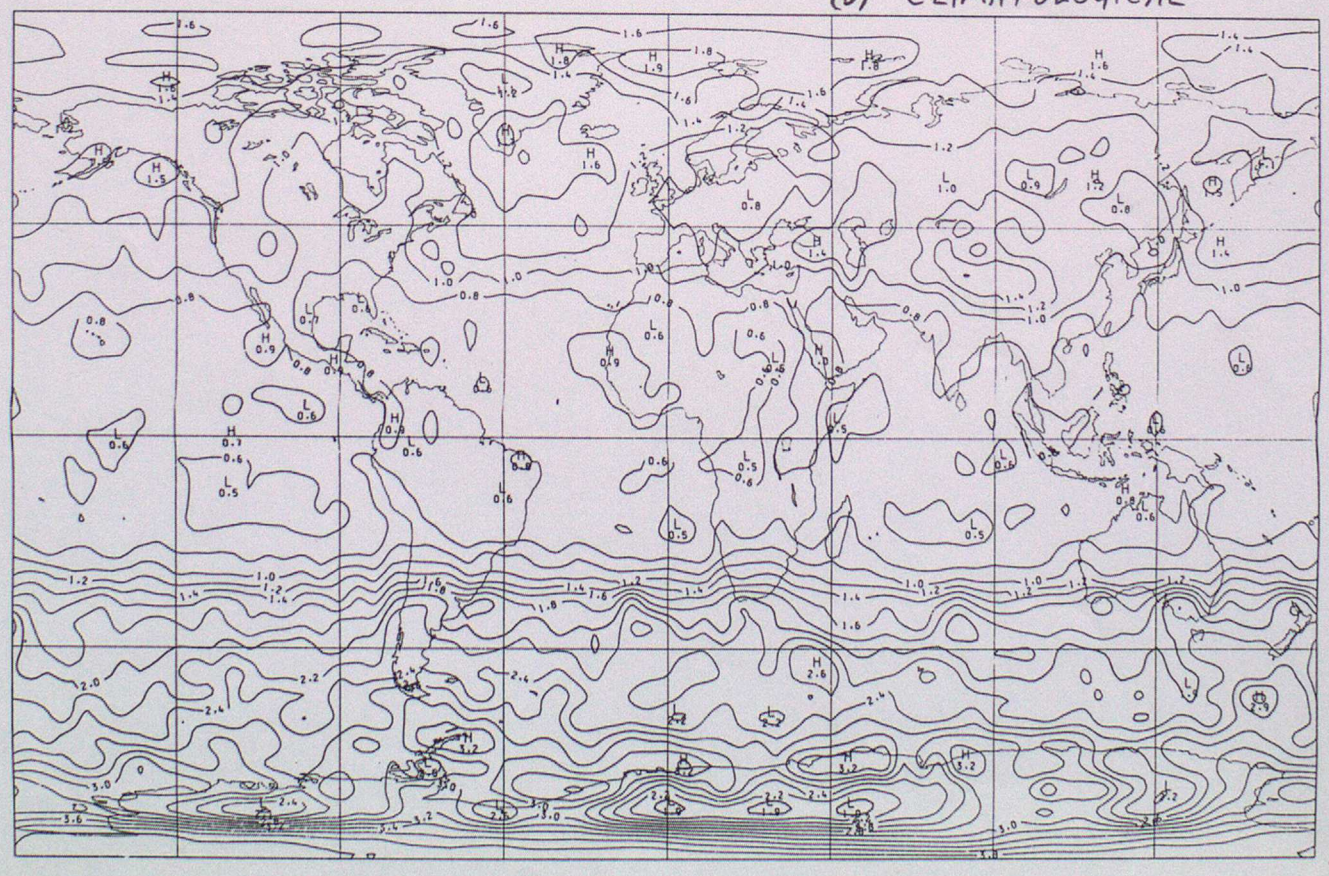


Figure 3 (a)

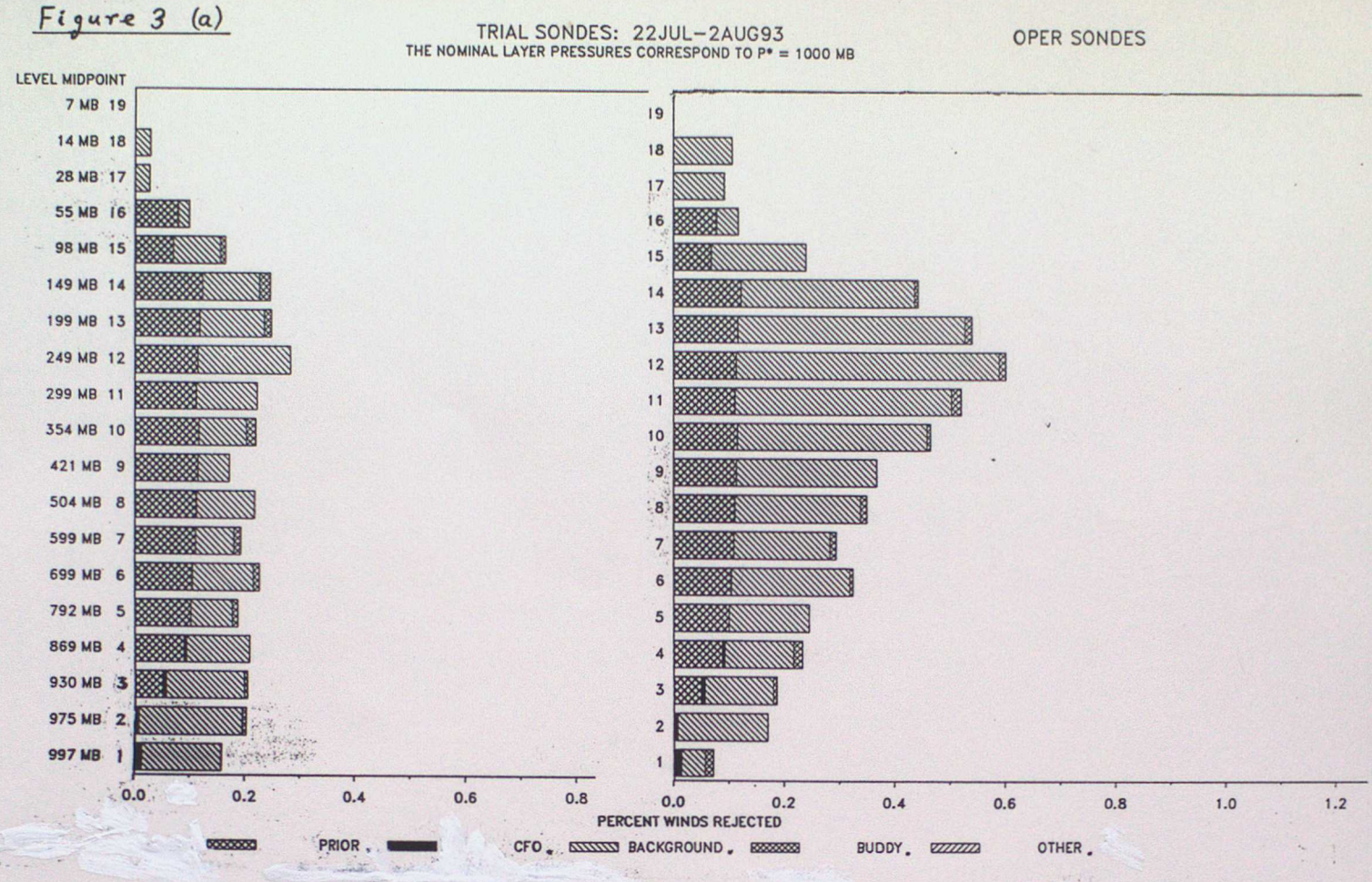


Figure 3 (b)

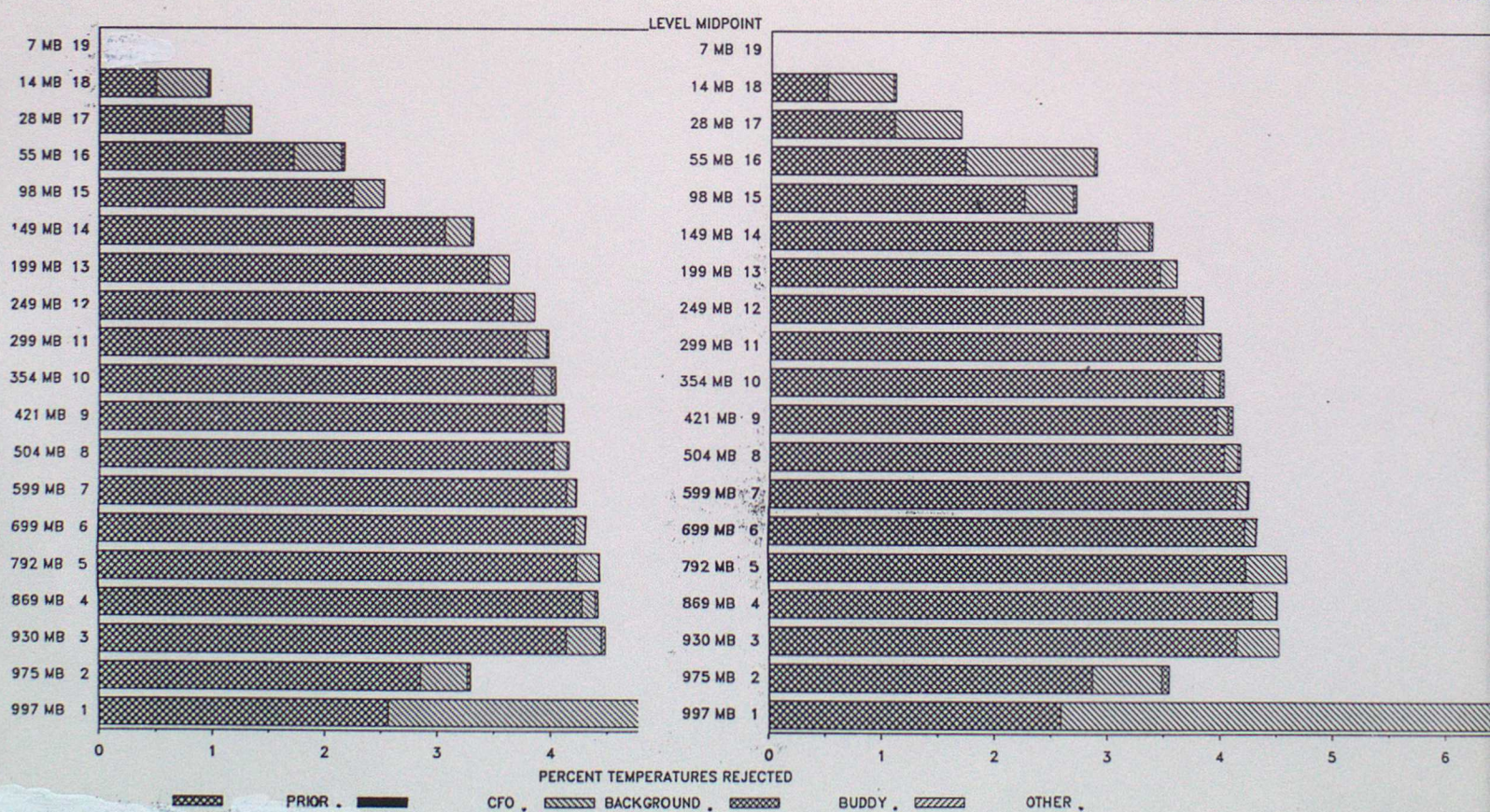
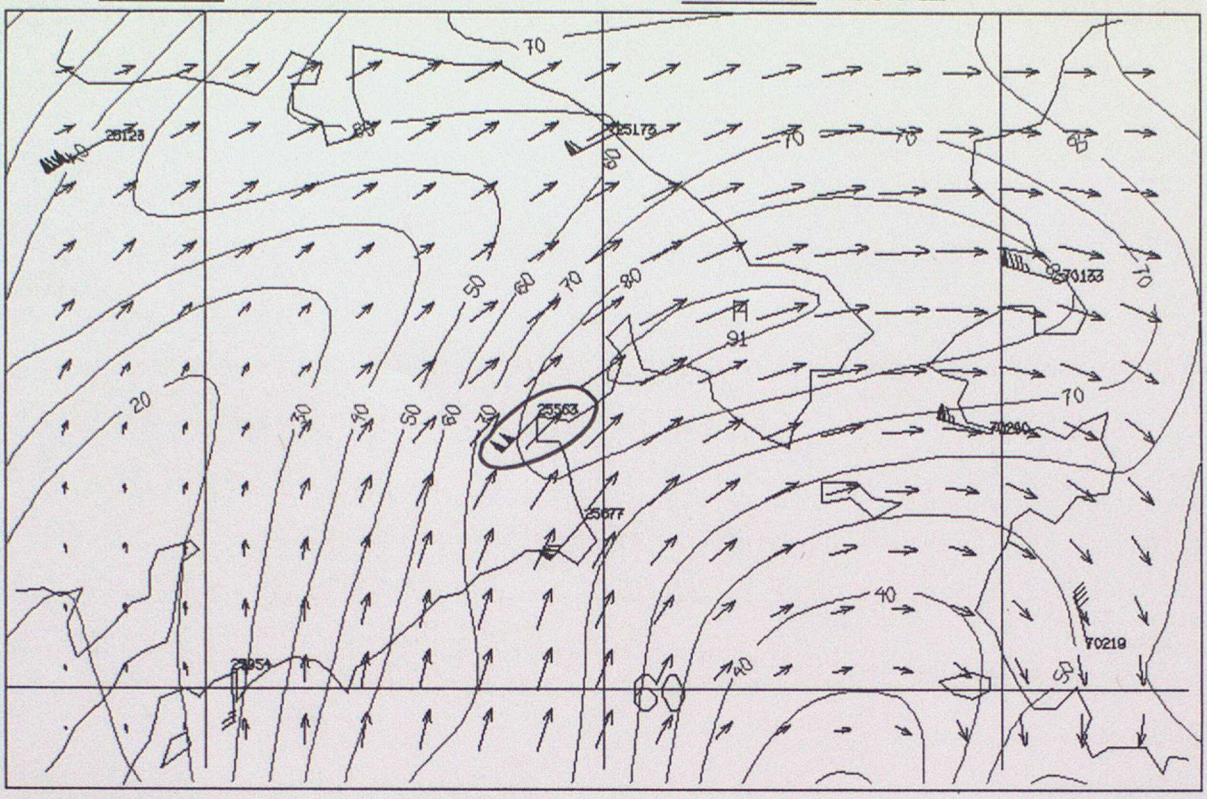


Figure 4 (a)

Ob flagged oper., passed qc in Trial.

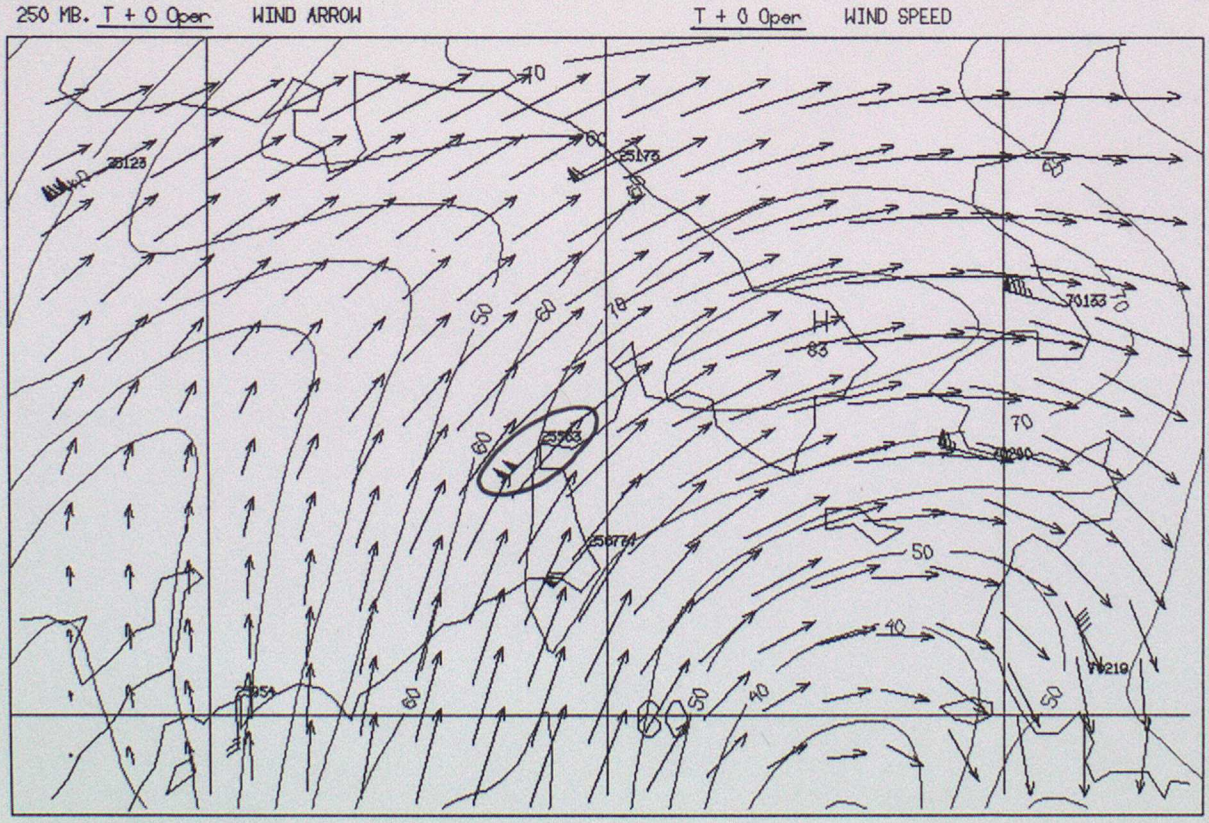
Oper OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
CALLSIGN T - 180 TO T + 180
12Z 1 / 8 / 1995. GRID CENTRE 65.0 180.0 SCALE: 8.00
250 MB. T + 0 Trial WIND ARROW T + 0 Trial WIND SPEED



→ REPRESENTS 50 KNOTS CONTOUR INTERVAL: 10 KNOTS 13 OBS 9 PLOTTED
0 B BOGUS 9 R TEMP 0 P PILOT 0 U TSHIP 0 A AIREP
0 M AMDAR 0 C ACARS

Figure 4 (b)

Oper OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
CALLSIGN T - 180 TO T + 180
12Z 1 / 8 / 1995. GRID CENTRE 65.0 180.0 SCALE: 8.00
250 MB. T + 0 Oper WIND ARROW T + 0 Oper WIND SPEED

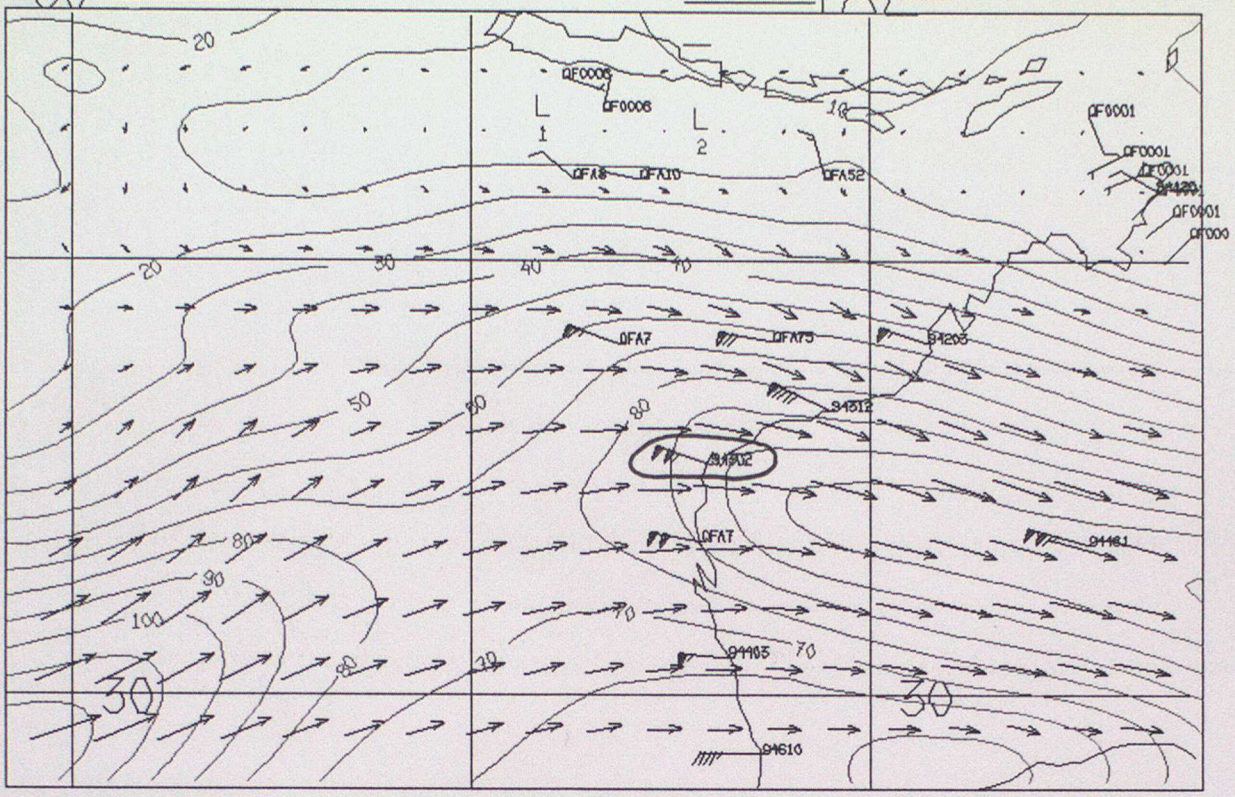


→ REPRESENTS 20 KNOTS CONTOUR INTERVAL: 10 KNOTS 13 OBS 9 PLOTTED
0 B BOGUS 9 R TEMP 0 P PILOT 0 U TSHIP 0 A AIREP
0 M AMDAR 0 C ACARS

Figure 5 (a)

Ob flagged oper, passed qc in Trial

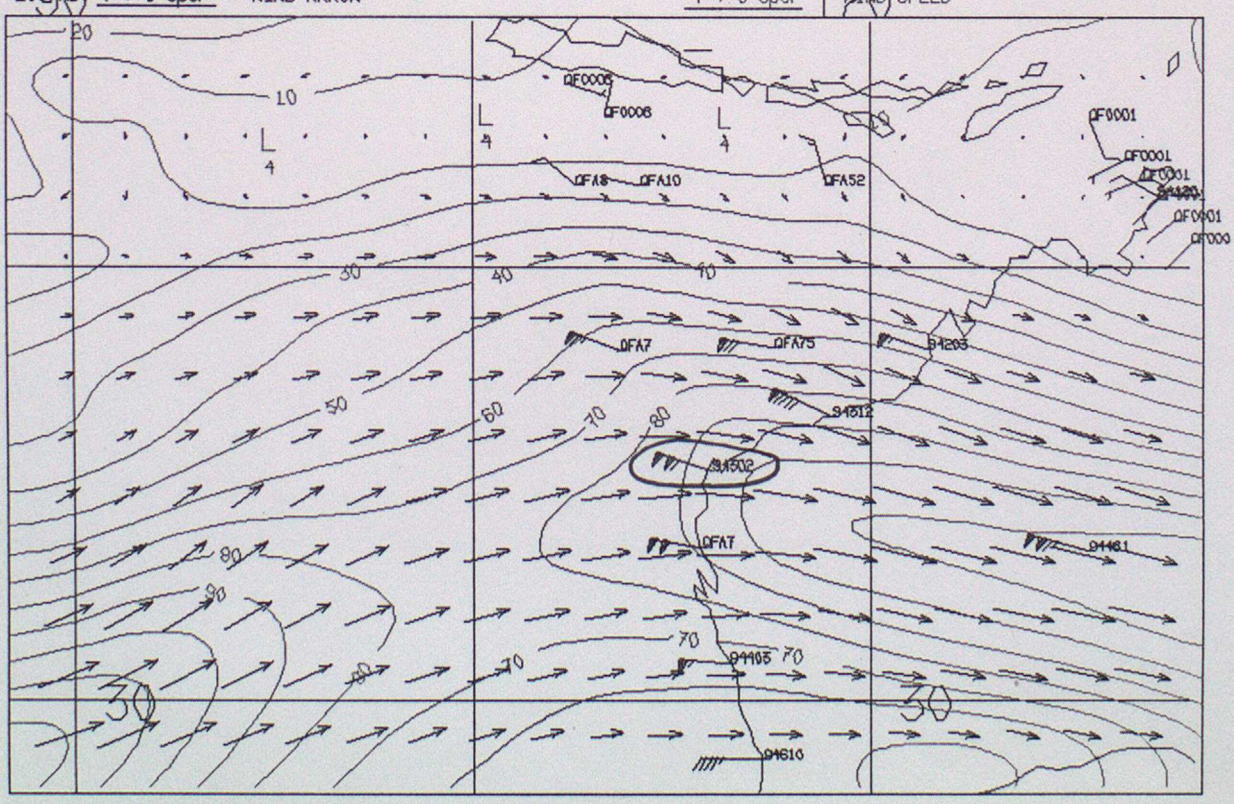
Oper OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
CALLSIGN T - 180 TO T + 180
12Z 1 / 8 / 1993. GRID CENTRE -20.0 110.0 SCALE: 8.00
250 MB T + 0 Trial WIND ARROW T + 0 Trial WIND SPEED



→ REPRESENTS 50 KNOTS CONTOUR INTERVAL: 10 KNOTS 37 OBS 21 PLOTTED
0 B BOGUS 0 R TEMP 7 P PILOT 0 U TSHIP 10 A AIREP
4 M AMDAR 0 C ACARS

Figure 5 (b)

Oper. OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
CALLSIGN T - 180 TO T + 180
12Z 1 / 8 / 1993. GRID CENTRE -20.0 110.0 SCALE: 8.00
250 MB T + 0 Oper WIND ARROW T + 0 Oper WIND SPEED

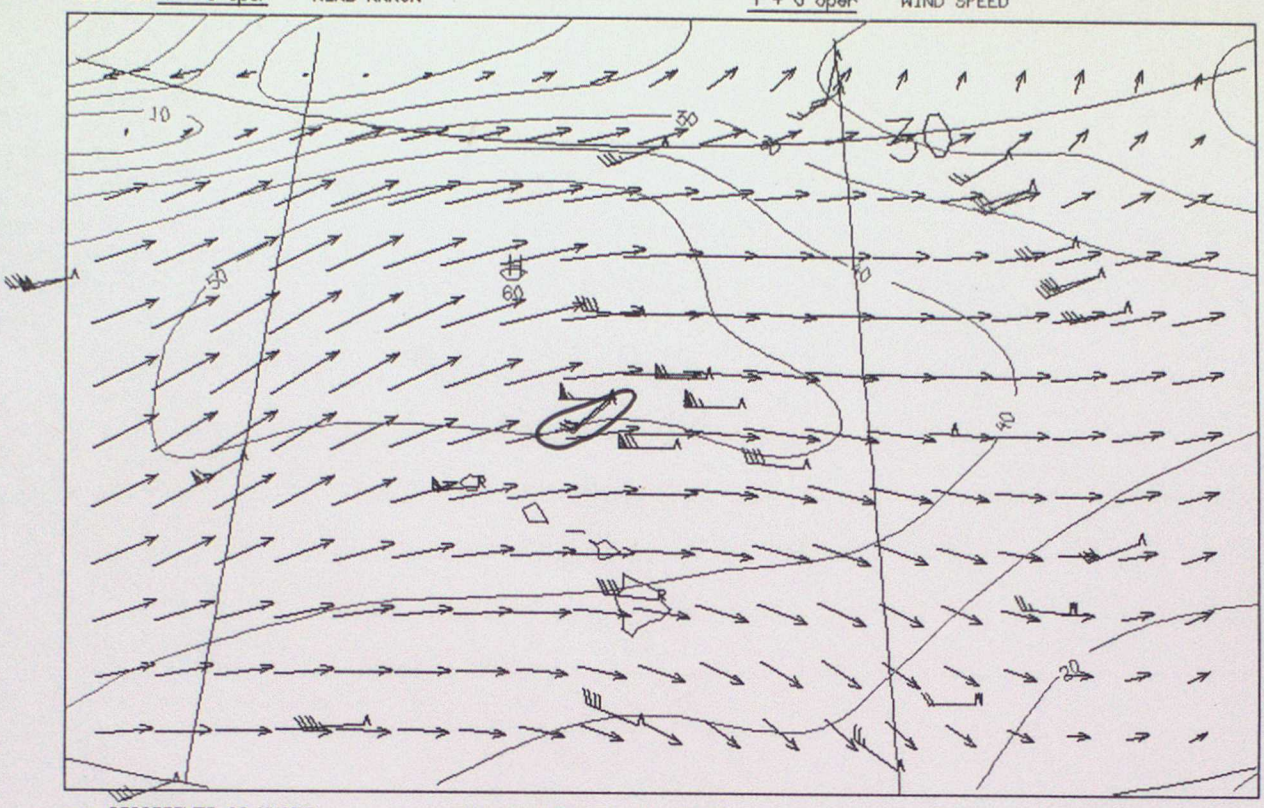


→ REPRESENTS 50 KNOTS CONTOUR INTERVAL: 10 KNOTS 37 OBS 21 PLOTTED
0 B BOGUS 0 R TEMP 7 P PILOT 0 U TSHIP 10 A AIREP
4 M AMDAR 0 C ACARS

Figure 6 (a)

ob flagged in Trial, passed qc oper.

Oper OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
OB TYPE T - 180 TO T + 180
12Z 31 / 7 / 1993. GRID CENTRE 24.0 -155.0 SCALE: 6.00
250 MB. T + 0 oper WIND ARROW T + 0 oper WIND SPEED

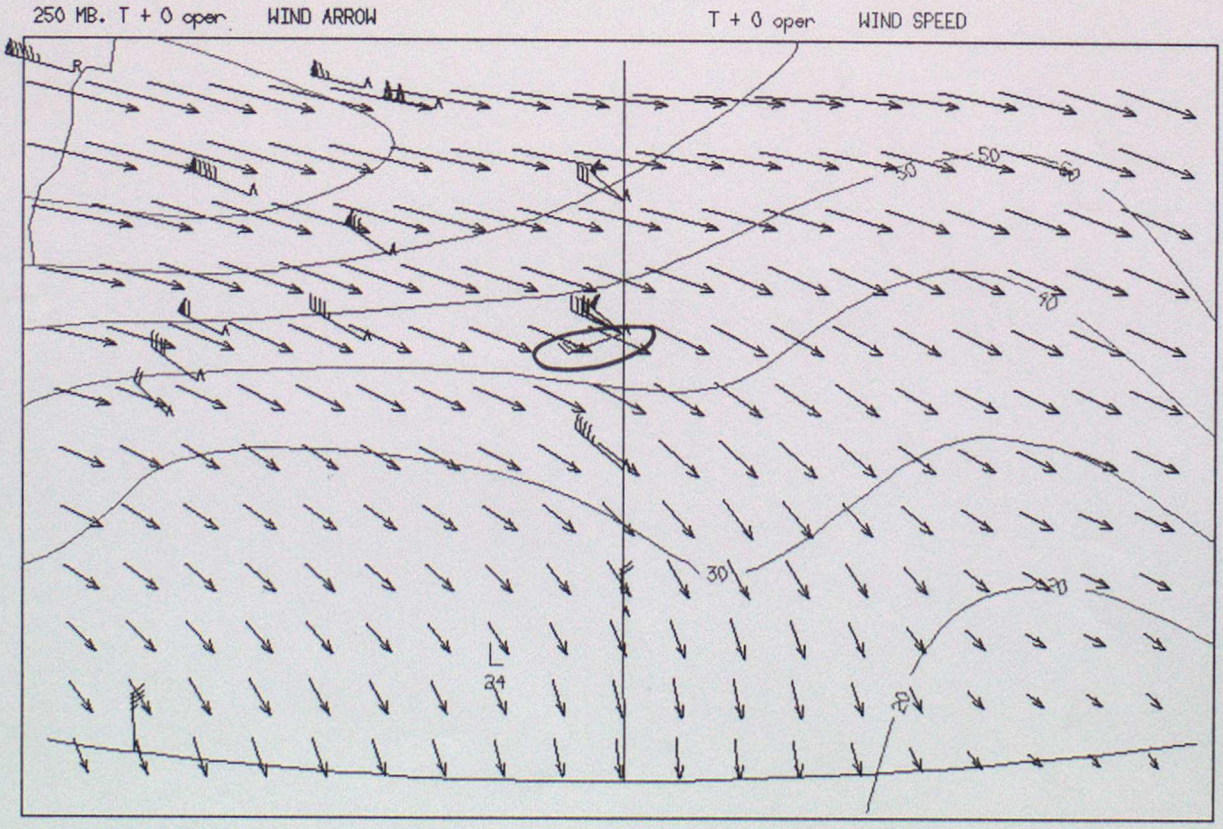


→ REPRESENTS 20 KNOTS CONTOUR INTERVAL: 10 KNOTS 59 OBS 38 PLOTTED
0 B BOGUS 2 R TEMP 0 P PILOT 0 U TSHIP 34 A AIREP
2 M AMDAR 0 C ACARS

Figure 6 (b)

ob flagged in Trial, passed qc oper.

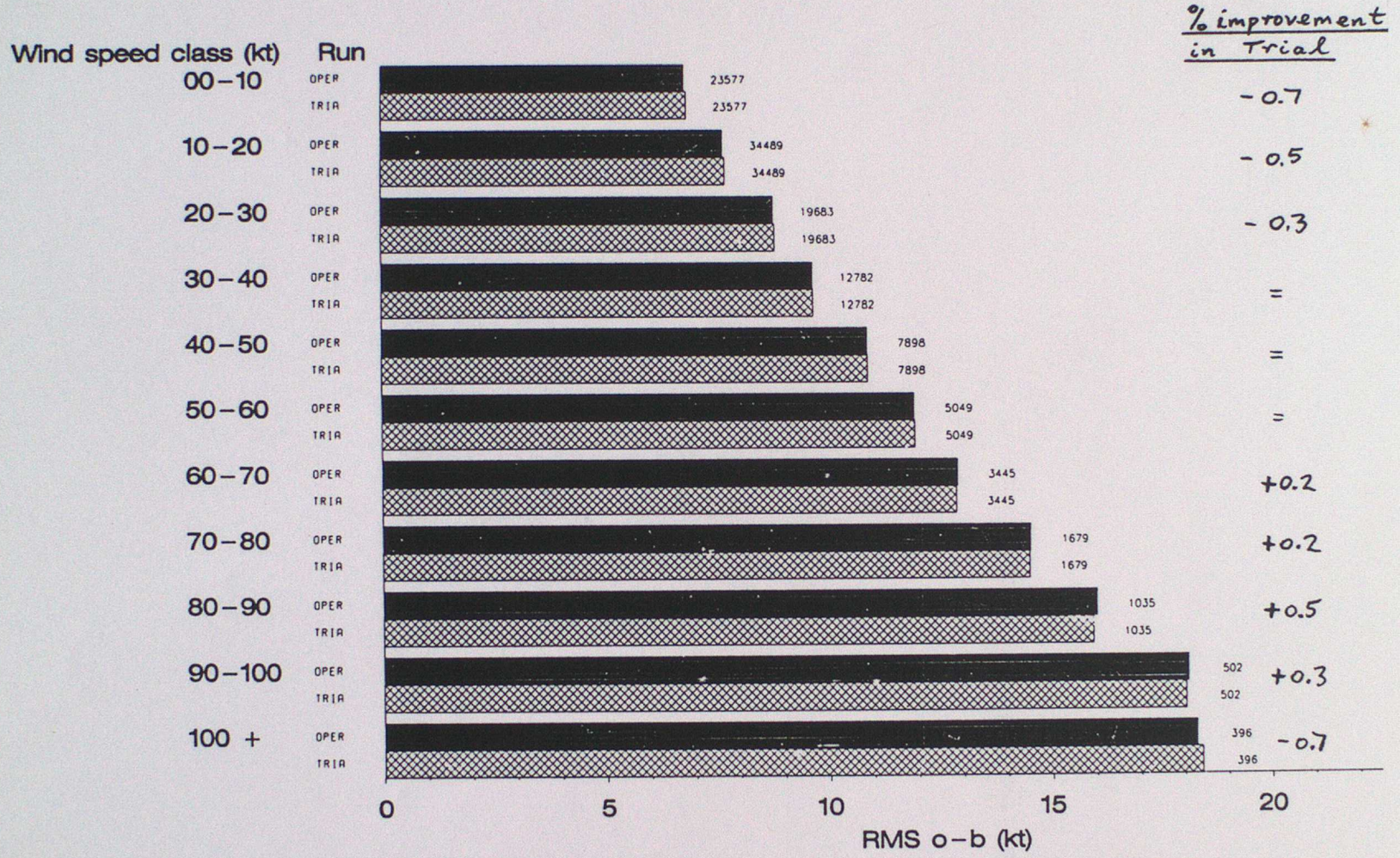
Oper OPD OBSERVATIONS 300.0 MB TO 200.0 MB.
OB TYPE T - 180 TO T + 180
12Z 31 / 7 / 1993. GRID CENTRE 34.3 149.9 SCALE: 12.00
250 MB. T + 0 oper WIND ARROW T + 0 oper WIND SPEED



→ REPRESENTS 20 KNOTS CONTOUR INTERVAL: 10 KNOTS 66 OBS 18 PLOTTED
0 B BOGUS 1 R TEMP 0 P PILOT 0 U TSHIP 17 A AIREP
0 M AMDAR 0 C ACARS

Figure 7.

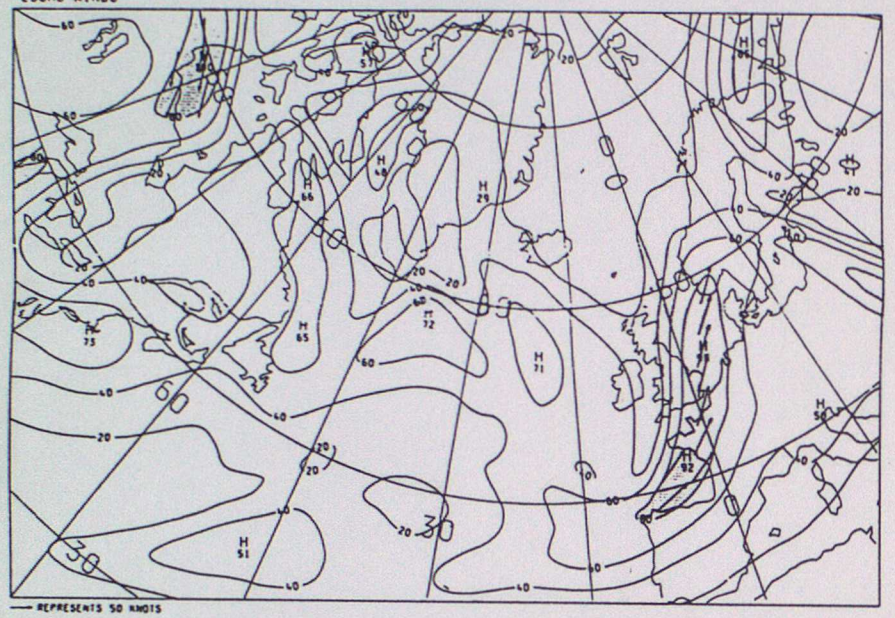
RMS vector wind differences (o−b) for Oper. and Trial runs
Radiosondes: Northern Hemisphere (30−90N)



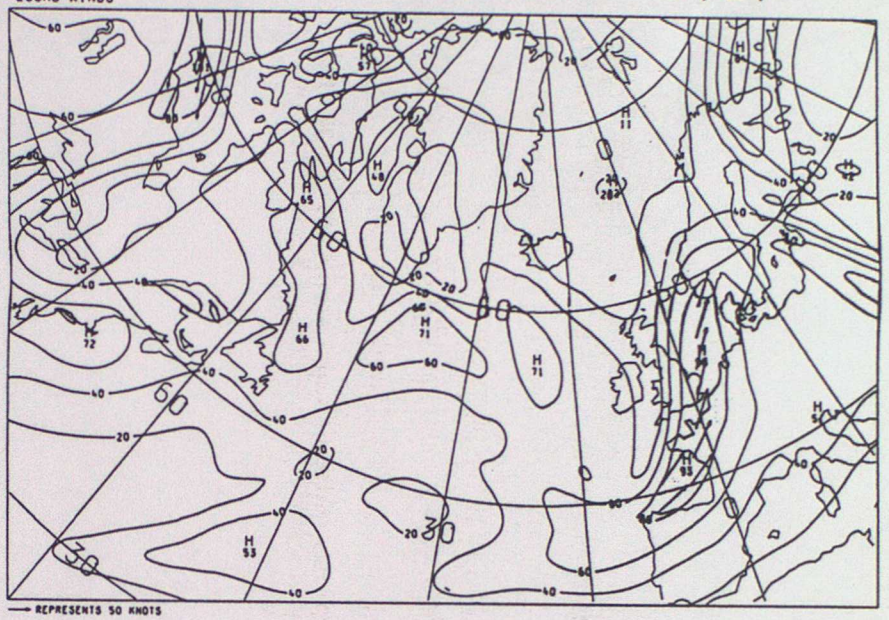
PERIOD: 22JUL-02AUG 1993
OBSERVATION SAMPLE: PASSED QC IN BOTH RUNS

Figure 8.

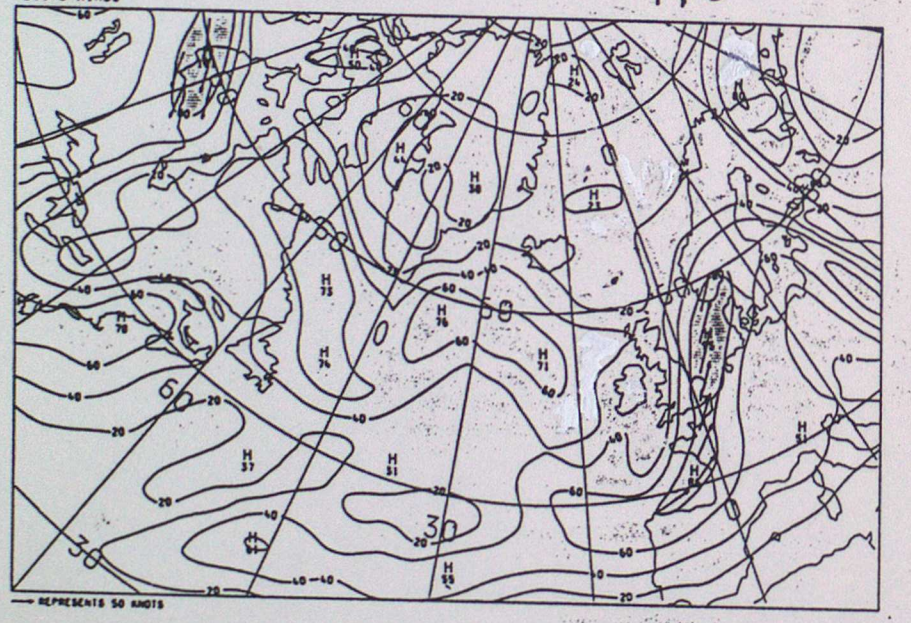
(a) Trial T+24



(b) Oper. T+24



(c) Verifying T+0



(d) Trial - Oper. T+24

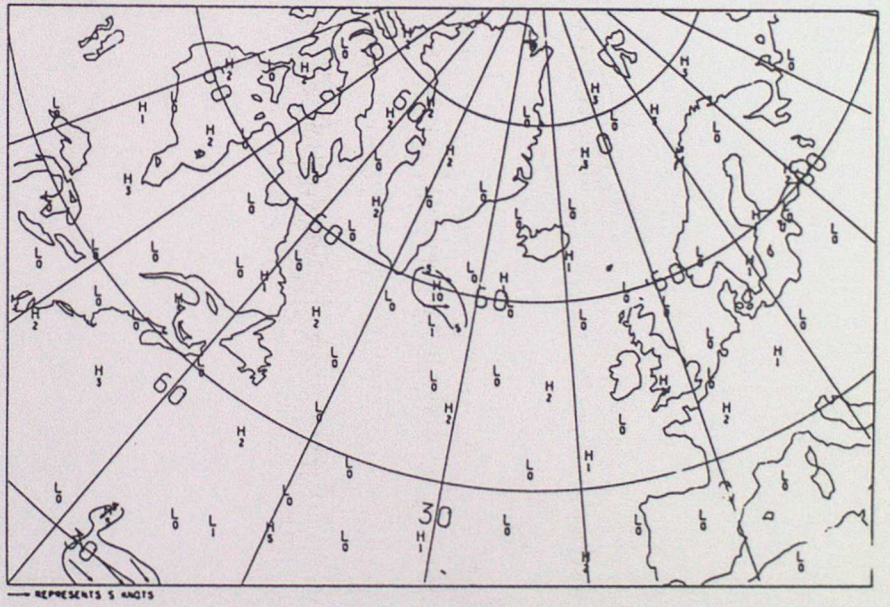
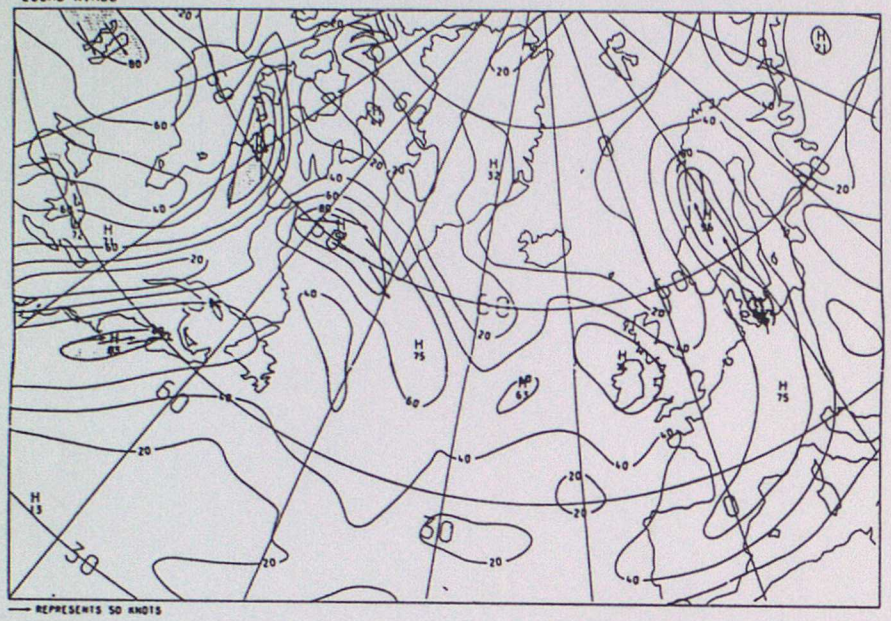
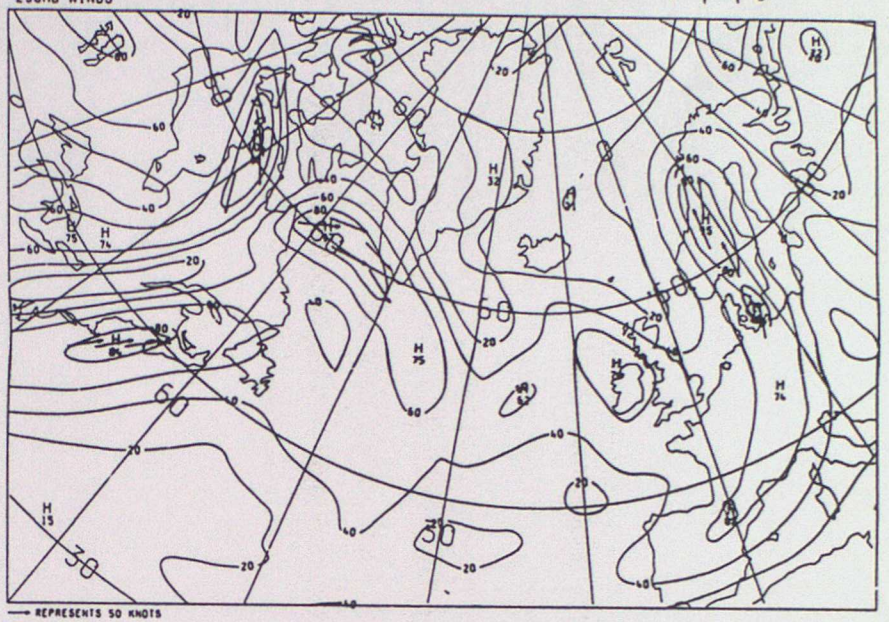


Figure 9.

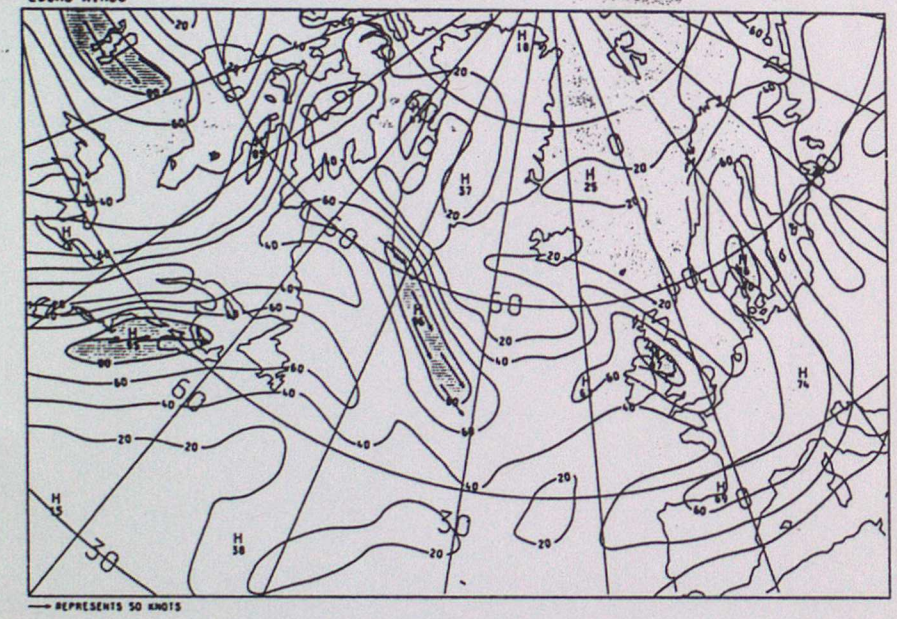
(a) Trial T+48



(b) Oper. T+48



(c) Verifying T+0



(d) Trial - Oper. T+48

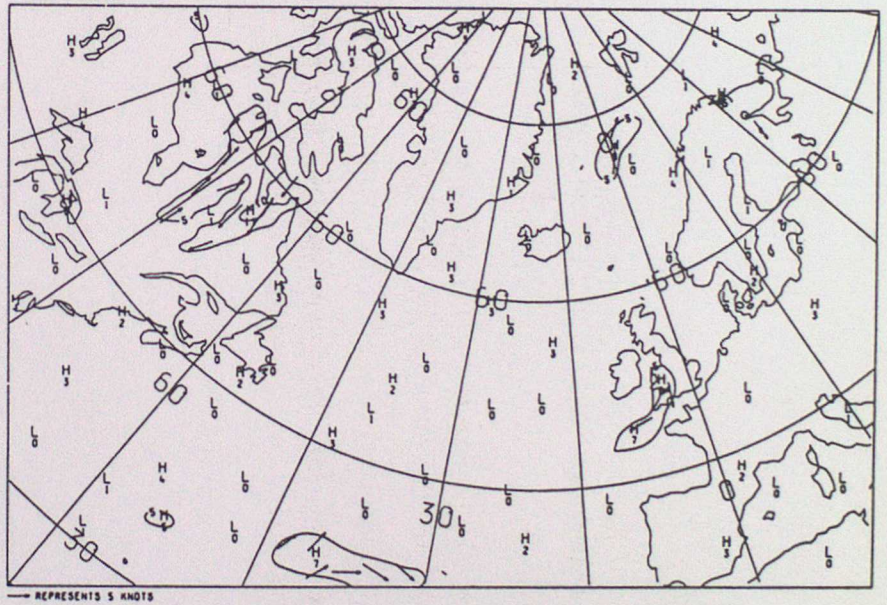
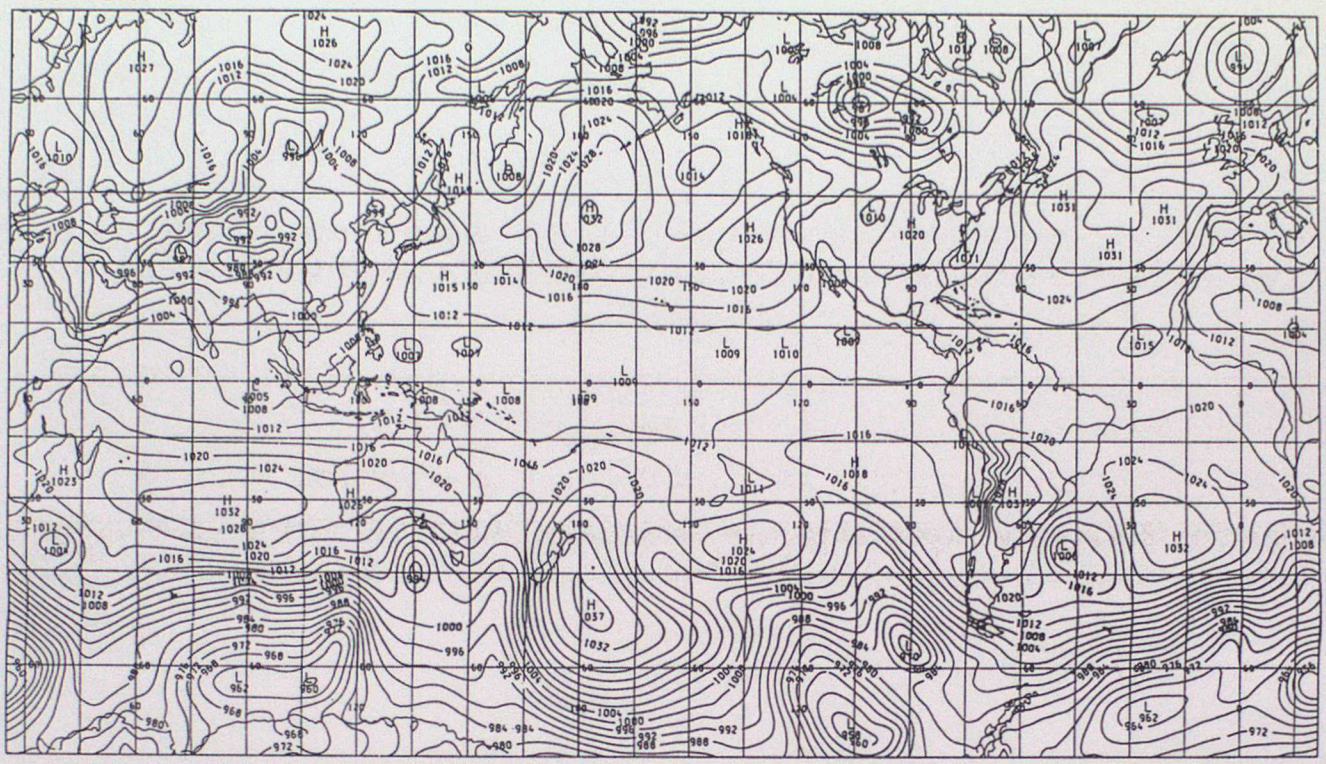


Figure 10.

(a) SYN-DEP BGE TRIAL - JUL93
TRIAL
VALID AT 12Z ON 31/7/1993 DATA TIME 12Z ON 28/7/1993 T+72
MSL PRESSURE



(b) SYN-DEP BGE TRIAL - JUL93
OPERATIONAL
VALID AT 12Z ON 31/7/1993 DATA TIME 12Z ON 28/7/1993 T+72
MSL PRESSURE

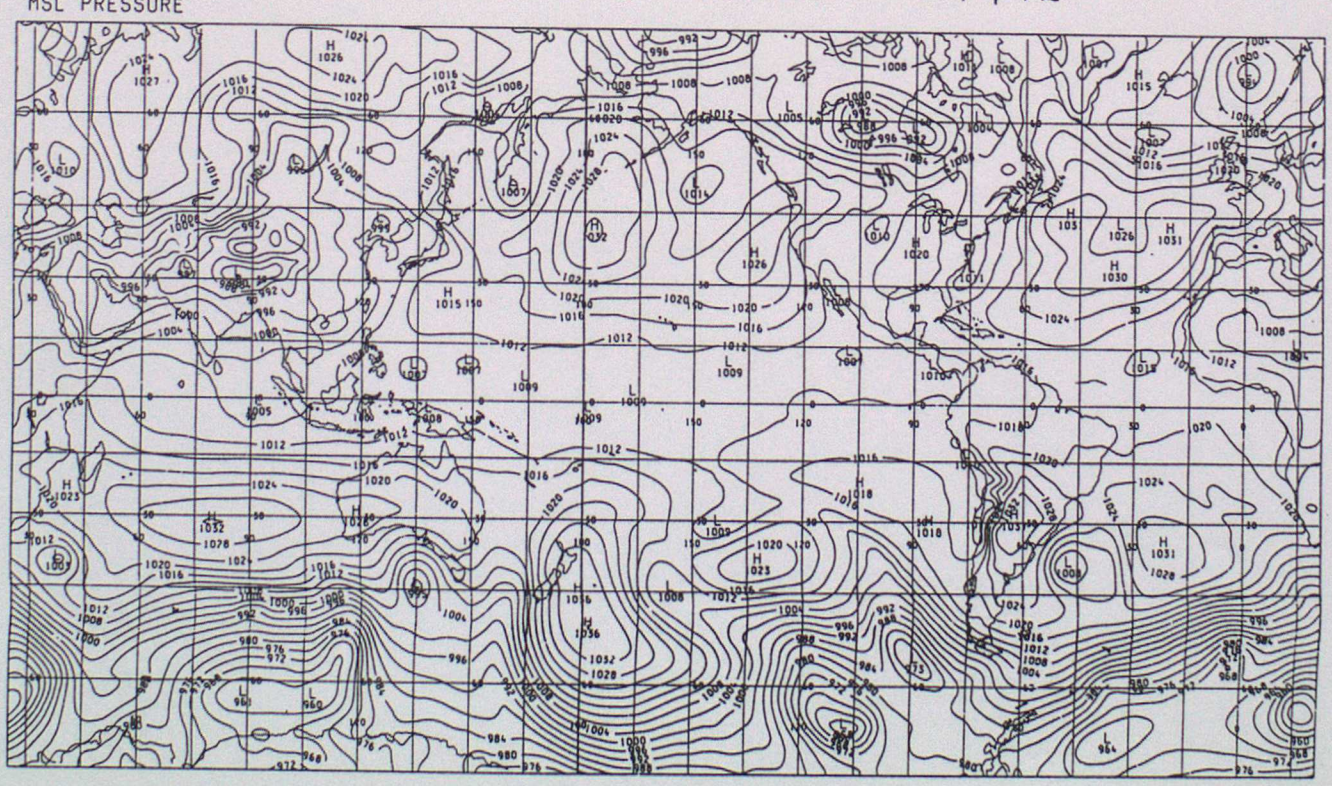
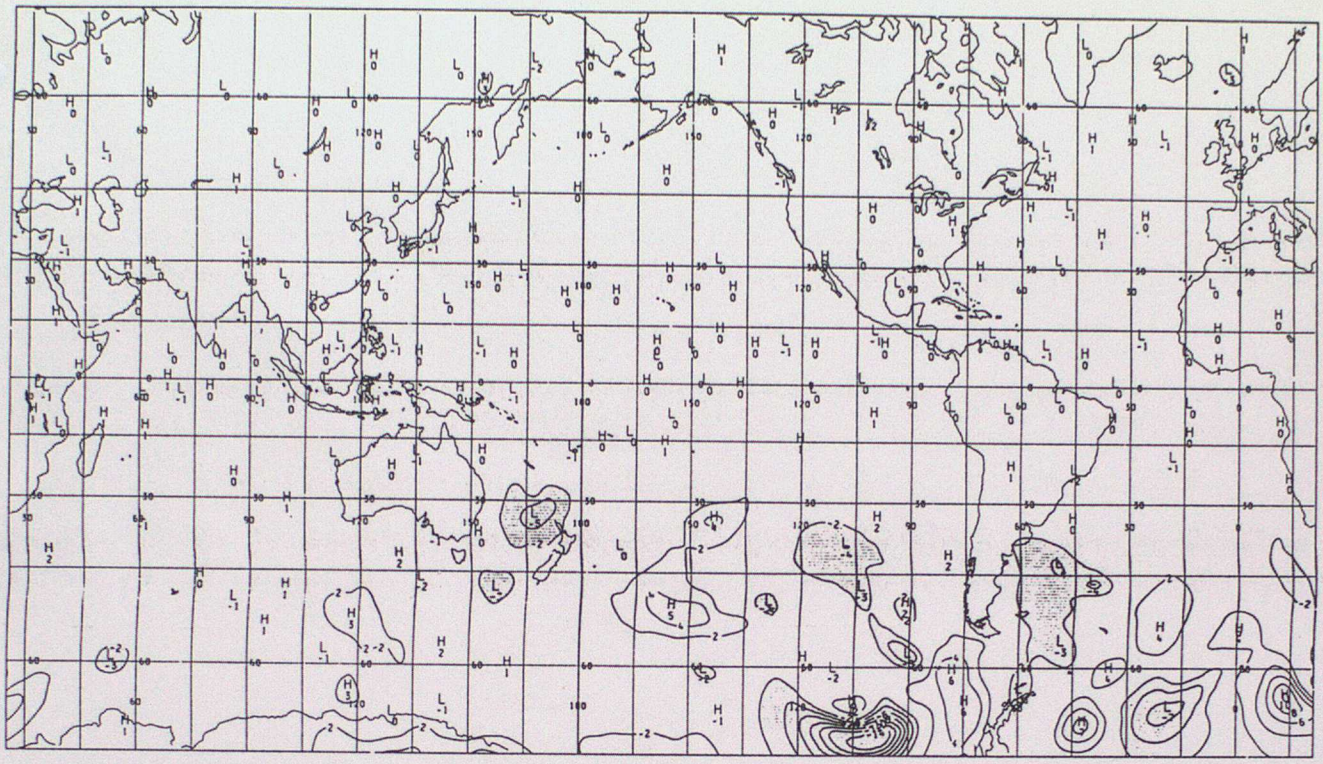


Figure 10. (cont.)

(c) SYN-DEP BGE TRIAL - JUL93
DIFF: Trial − Oper
VALID AT 12Z ON 31/7/1993 DATA TIME 12Z ON 28/7/1993 T+72
MSL PRESSURE



(d) SYN-DEP BGE TRIAL - JUL93
OPERATIONAL
VALID AT 12Z ON 31/7/1993 DATA TIME 12Z ON 31/7/1993 Verifying Anal. T+0
MSL PRESSURE

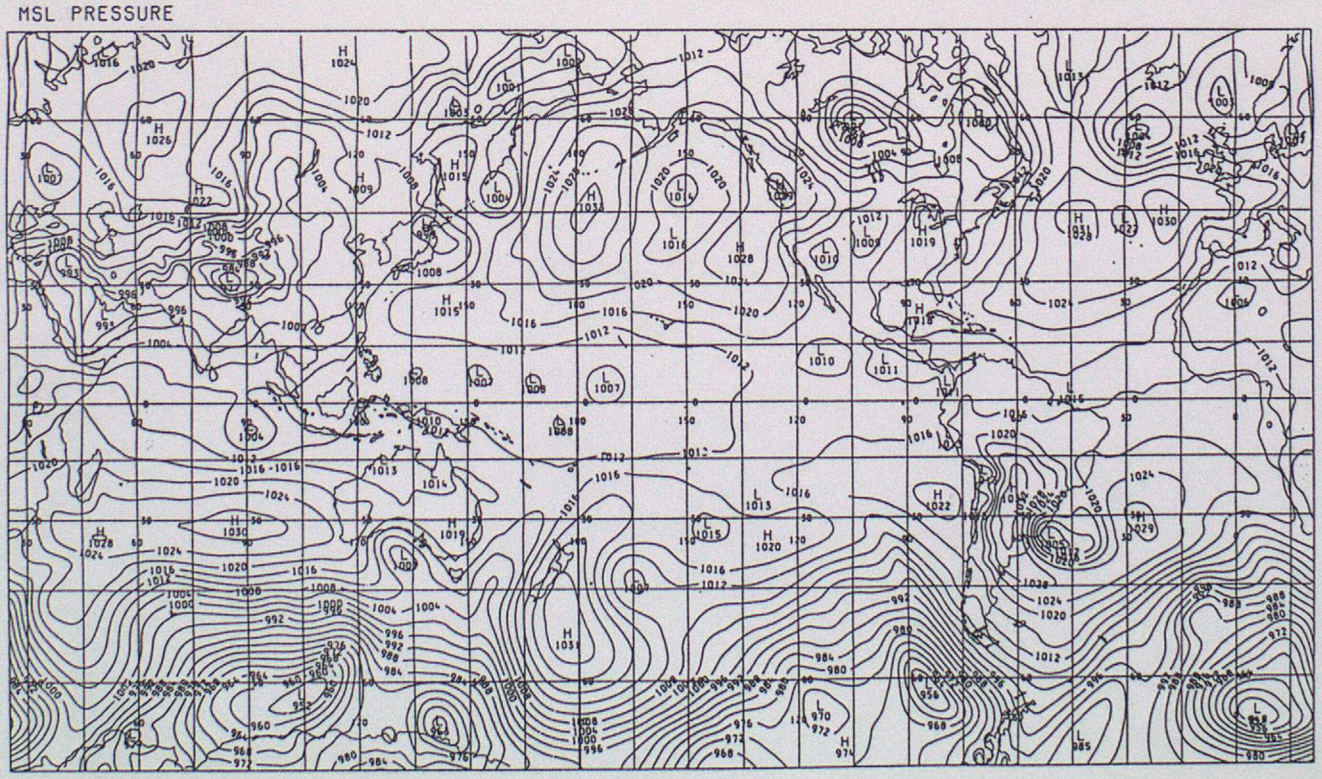


Figure 11.

SYN-DEP BGE TRIAL - JUL93
(a) THICKNESS DIFF: Trial - Oper
VALID AT 12Z ON 1/8/1993 DATA TIME 12Z ON 30/7/1993 T+48
1000-500 TT AND 500 HT

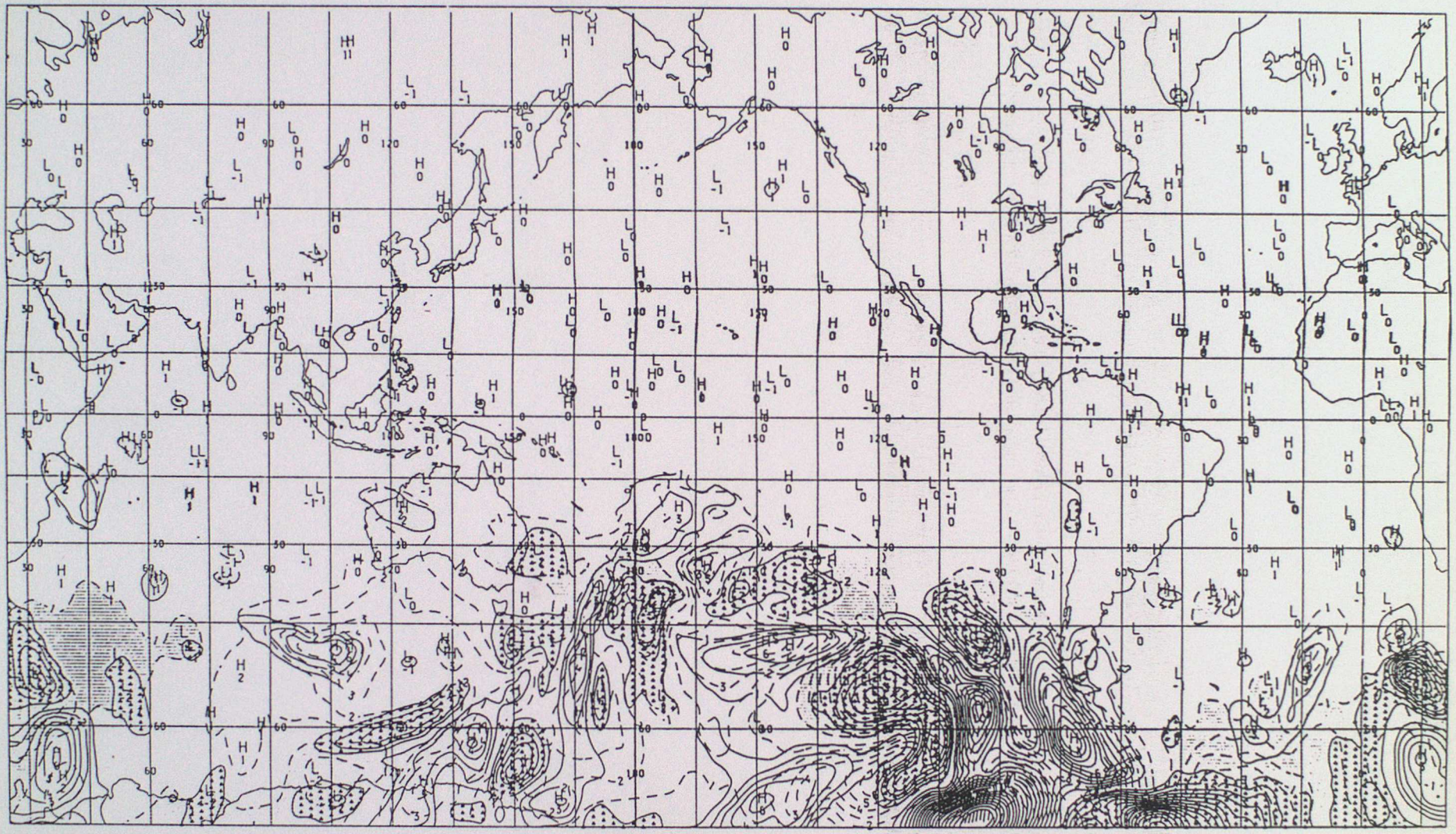
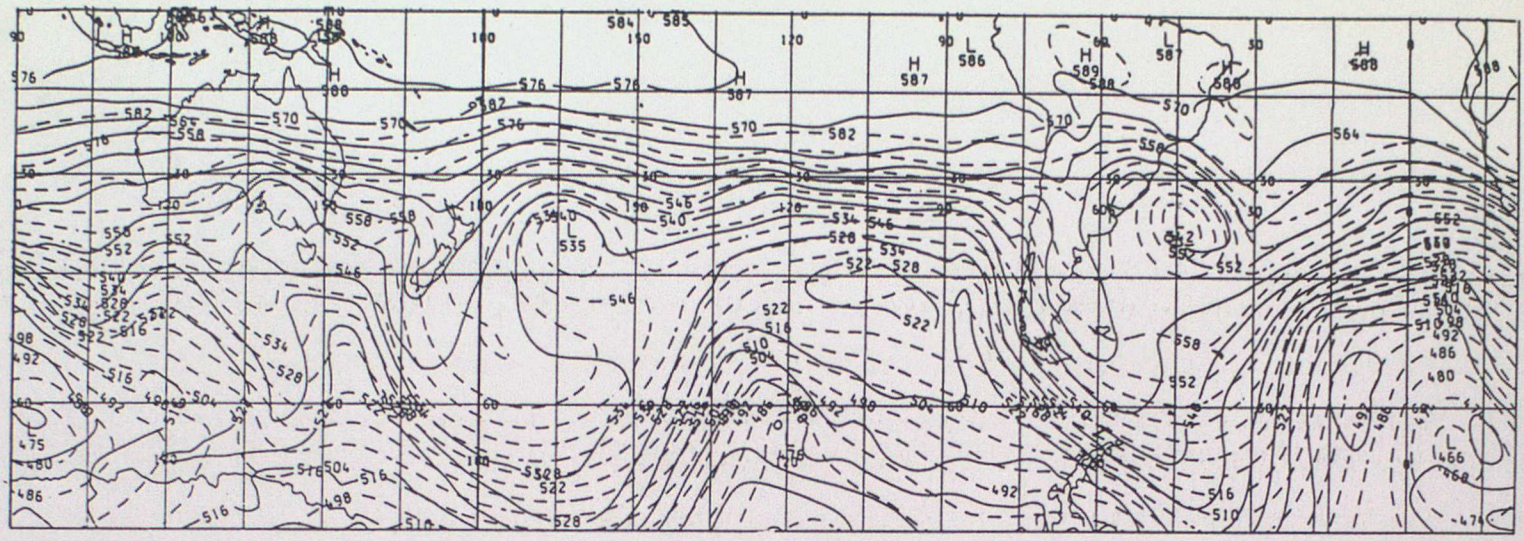


Figure 11. (cont.)

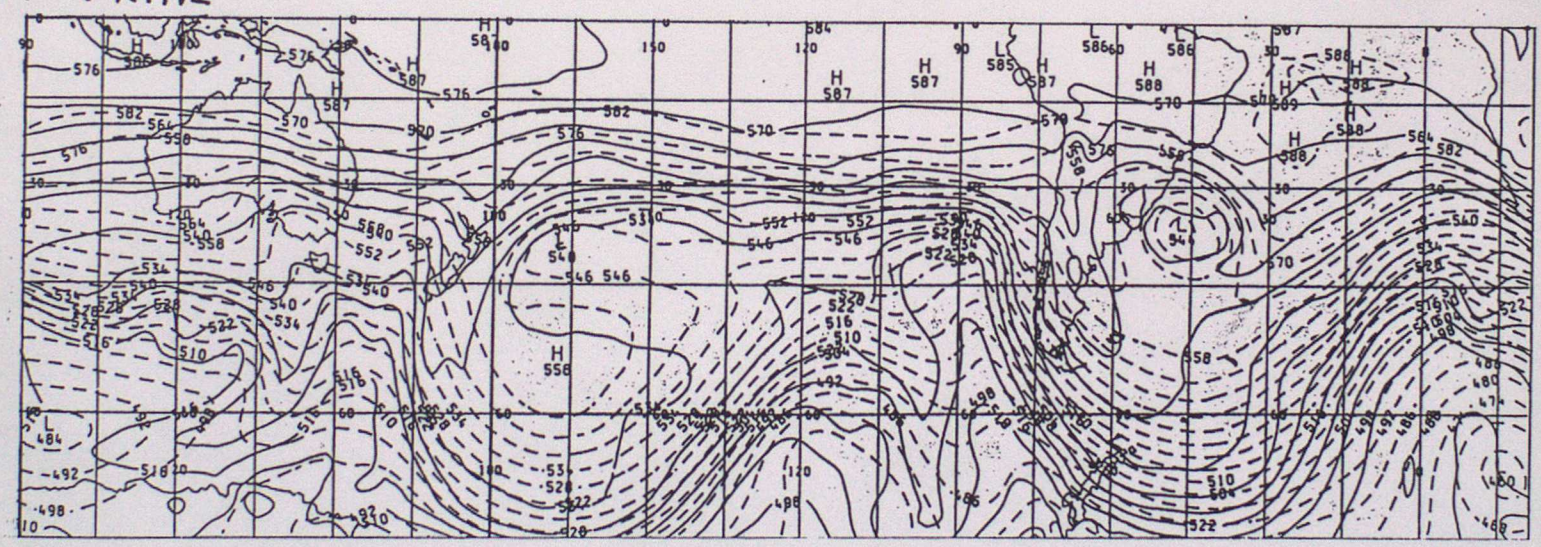
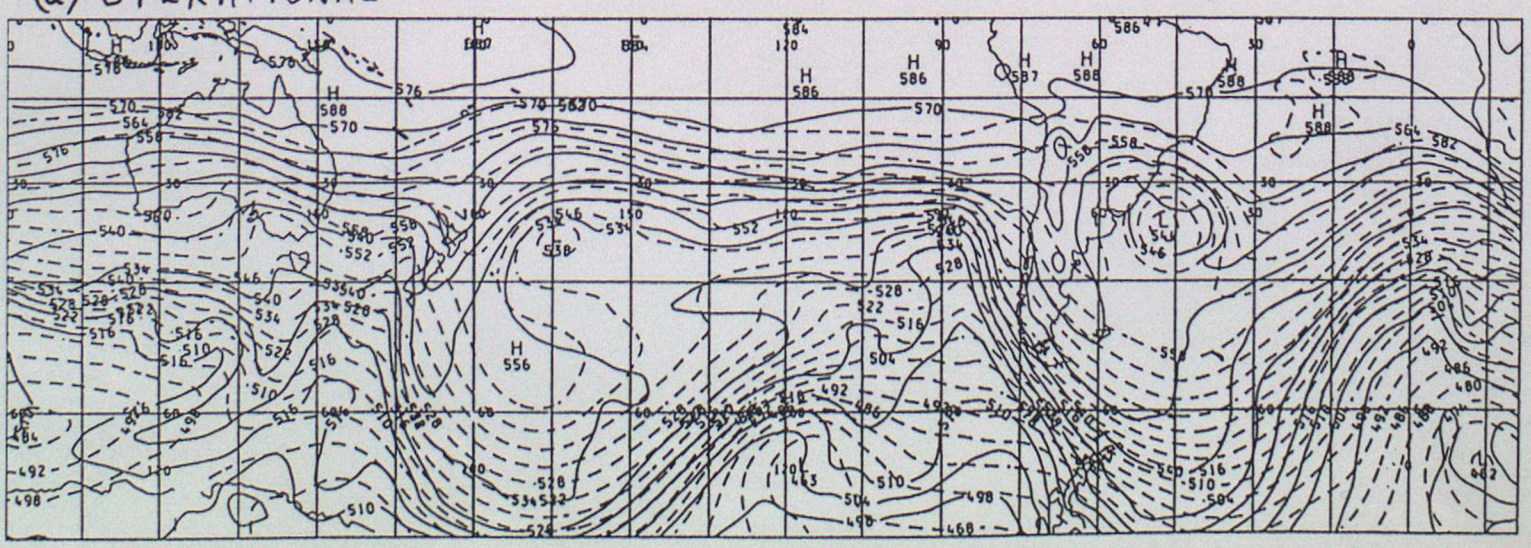
(b) SYN-DEP BGE TRIAL - JUL93
OPERATIONAL
VALID AT 12Z ON 1/8/1993
1000-500 TT AND 500 HT

Verifying Analysis
T+0



VALID AT 12Z ON 1/8/1993
DATA TIME 12Z ON 30/7/1993

(c) TRIAL T+48



(d) OPERATIONAL T+48